

THE EVOLUTIONARY LEARNING ALGORITHM IN PROCESSING OF ECOLOGICAL DATA, CASE-STUDY : POPLAR TREE-RING ANALYSIS

Hans Beeckman¹ & Kristine Vander Mijnsbrugge²

SUMMARY

- (1) Tree-rings are a challenge for modern ecological research dealing with 'time' and 'complexity'.
- (2) Machine learning appears to be an efficient tool to extract knowledge from data sets showing a high degree of complexity.
- (3) The software package PC/BEAGLE, using the evolutionary learning algorithm, is able to find some high quality rules for Nelder design poplar plantations.
- (4) Poplar trees growing in thin stands show a strong negative temperature signal.
- (5) 'Beaupré'-trees are almost more massive than 'Unal'-trees, except for the first years after planting in the wider planting distances.

INTRODUCTION

'Time' and 'complexity', two concepts often denied by the scientific enterprise in its classical definitions, are a main challenge for modern science (Prigogyne & Stengers, 1985). Most of the ecosystems are a 'messy laboratory' for researchers. They have to take into account a lot of variables and an extraordinary quantity of interactions which are difficult to survey and which can be situated at different levels of system integration. Identification of the distinct environmental factors and the individualization of the responses of a population to each factor are not always possible. Moreover, some processes, such as the delays in the effect of density dependence, lead to complex responses that have no simple relation with the changes in the environment.

Anyhow, such relations have always been very important for ecological science. Ecological studies assume implicitly that there is a network of interactions and feedbacks connecting all species in a pattern which is consistent in time and space: the ecosystem. 'Ecosystem structure and function' is understood to refer to those aspects which appear only at the level of the ecosystem, rather than its components. Ecologists focus in particular into those biological interactions and regulatory feedbacks which tie the system together and cause it to behave as a recognizable integrated system, rather than as a collection of independent populations (Odum, 1971).

In the early stage of ecological research, it was clear that ecologists could only make use of totally deterministic models in rather rare circumstances where populations are small enough to localize and measure all population elements. In addition, most of the relations are not monothetic : a huge quantity of species is involved. Much more relations are uni- or multimodal,

* Laboratory of Plant Ecology, State University of Ghent, Coupure links 653, B-9000 Gent, BELGIUM.

** BIKIT (Babbage Institute For Knowledge and Information Technology) associated with State University of Ghent, Joseph Plateaustraat 22, B-9000 Gent, BELGIUM.

also showing optima and limits, than linear. The triangle axiom is not always fulfilled, what makes the Euclidean space not useful as a geometric model in certain cases. Causal relations are difficult to detect, all the more because in ecosystems network causalities are more frequent than linear chain causalities of one cause and one effect.

Through all these difficulties, the research stages of explaining and interpreting are often never reached: the knowledge acquisition bottleneck is very small in ecology and many knowledge potion bottles stay unopened. Nevertheless there is no shortage of interesting data. 'Nothing is so detrimental to the acquisition of insight than an underestimation of the knowledge and reasoning in a particular domain' as Breuker and Wielenga (1983) put it.

Tree-rings for instance contain information about the total life span of a tree, thanks to the yearly formed wood increment. As such, they are valuable ecological sensors detecting environmental change on a time scale which is sometimes as long as several hundreds of years. Tree-ring analysis in which 'time' and 'complexity' are involved, has placed heavy demands on methodological improvement, concerning as well variable selection and methods of measurement as in data processing.

The assessment of tree-ring-environment relations is hampered by the complexity of forest ecosystems. The response of trees to weather for instance might depend on the planting distance, which causes extreme differences in growing conditions, as well as the response of trees might change according to genetic code of the tree or according to the tree age.

To evaluate the meteorological signal as a function of stand density and stand age, a man-made poplar ecosystem, which is genetic and edaphic homogenous, is investigated.

The question arises if Artificial Intelligence methodologies can be useful in processing of complex data sets, consisting of tree-ring variables and external explanatory variables.

MATERIAL AND METHODS

Tree-rings from a Nelder design poplar plantation

(i) Clones

Poplar and poplar plantations have always been a characteristic feature of the Flemish landscape. Poplar is also important to the total wood production in Belgium : about 15 % of the annual production is poplar, while poplar stands contribute only 6.5 % to the total forested area (Schalck, 1982).

Populus x euramericana (Dode) Guinier 'Unal' and 'Beaupré' are two of the famous fast growing hybrid clones of Populus trichocarpa and Populus deltoides, realized by extensive breeding work in 1961 by the State Equipment Station for Poplar Culture, Geraardsbergen, Belgium (d'Oultremont and Steenackers, 1973). These poplar clones show extremely good productivity, growth characteristics and disease resistance.

(ii) Experimental fields

The first experimental field is part of the Provinciaal Domein Puyenbroeck at Wachtebeke (Oost-Vlaanderen), with geographical coordinates latitude 51°09'25''N and longitude 3°54'00''E, constitutes a large recreational area of about 400 hectares

(Fig.1). The area is planted with poplar trees using row-type and lane-type planting patterns. A rectangular field of 7.81 ha has been reserved for field experiments in relation to meteorological, hydrological and primary production programs. A first part of the field was planted in spring 1974 with two poplar clones 'Unal' and 'Beaupré' (Lemeur *et al.*, 1976). A second experimental poplar plantation is located at Kaulille (Bocholt-Limburg), being part of Lozerbos State Forest ($52^{\circ}12'30''N$; $5^{\circ}32'40''E$).

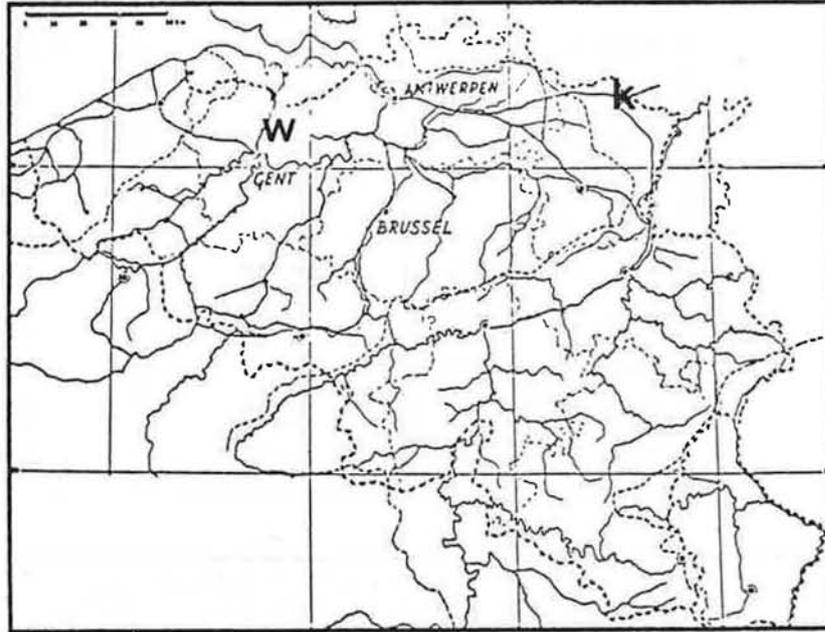


Figure 1 : Location of experimental fields.

(iii) Planting pattern

The planting pattern of the Wachtebeke experimental field was based on Nelder's (1962) design of concentric circles (Fig.2): the interplant distance on the circles increases with increasing distance from the midpoint (from 1.5 m on the inner to 8 m on the outer circle). The planting pattern permits to evaluate variable planting distances within a minimum of space (Table 1). The wide range of interplant distances provided rather extreme differences

in growing conditions.

A similar planting pattern has been used at Kaulille: thirteen concentric circles and a planting distance varying from 1.5 m on the inner circle to 9.17 m on the outer circle. The first growing season after planting is 1980.

(iv) Sampling

In spring 1990 15 'Beaupré' trees from different circles of the Wachtebeke plantation and hence interplant distances were cut for dendro-ecological and woodtechnological research. Transversal cross-sections taken at 1.30 m height were transported to the laboratory for dendrochronological investigation. In Kaulille stem diameter at breast height was measured, almost every year.

Table 1 : Sample trees and corresponding planting distance (m)

	circle	plant. dist.(m)	trees per ha
tree 1	2	1.71	3,943
tree 2	4	2.05	2,743
tree 3	6	2.46	1,908
tree 4	7	2.69	1,592
tree 5	9	3.23	1,107
tree 6	8	2.95	1,327
tree 7	10	3.54	923
tree 8	12	4.24	642
tree 9	12	4.24	642
tree 10	12	4.24	642
tree 11	14	5.08	447
tree 12	16	6.09	311
tree 13	16	6.09	311
tree 14	16	6.09	311
tree 15	17	6.67	259

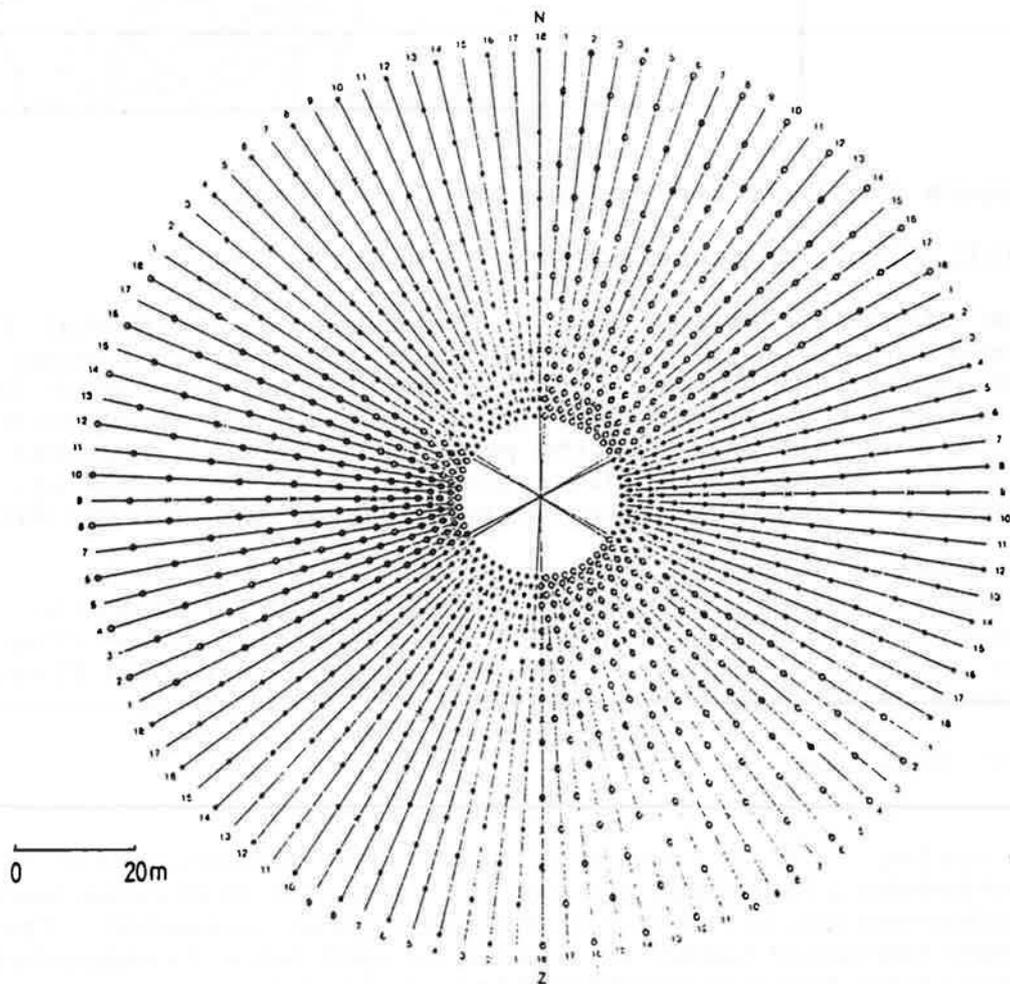


Figure 2 : The Nelder design planting pattern of the Wachtebeke experimental field.

(v) Data set

A first data file was constructed with ring widths and meteorological data, concerning the Wachtebeke field. The tree-ring widths were measured by a semi-automated image analysis system in four directions (N, E, S, W). The external explanatory variables are the april-september meteorological variables, calculated from the daily information on the weather from Melle, being a main meteorological station at 20 km from the experimental field. Seven meteorological variables are taken into account : (a) mean temperature ($^{\circ}\text{C}$), (b) amount of precipitation (mm), (c) mean relative humidity, (d) amount of sunshine (minutes), (e) mean percentage sunshine, (f) mean insolation (J/cm^2), (g) mean evaporation (mm). All variables are standardized by subtracting the mean and dividing by the standard deviation. This way there are some zero or negative cases as well as some positive ones. This is an internal housekeeping requirement which will make it possible for the BEAGLE program to handle numeric target expressions. An associated tag-file contains the variable names, indicating whether they are numeric or character data. Another data file was constructed with the stem diameters of the Kaulille plot from 1981 to 1989.

The evolutionary learning algorithm and data processing

(i) Artificial Intelligence

Artificial Intelligence is the part of computer science concerned with designing "intelligent computer systems", that is, systems that exhibit the characteristics we associate with intelligence in human behavior - understanding language, learning, reasoning, solving problems, uncertainty and so on (Boden, 1977). Or in other words, the field of AI is trying to produce both software and hardware that can simulate the characteristics of a human being which we classify as intelligent: learning, reasoning, problem solving. The development of AI brought with it the extension of the concept of "data" to "information", the latter being relevant data with respect to a clearly defined domain. AI also gave birth to the concept of "knowledge", being the ability to distinguish between data and information. Knowledge is concerned about objects, relationships, facts, rules,... Men or machines are intelligent to the extent to which they control and use knowledge.

At the heart of all expertise whether human or computerized, and hence of ecological research, lies pattern recognition (Fig. 3.) (Forsyth & Rada, 1986). In all pattern recognition tasks, there are a number of measurements made of an event or object. These raw measurements are transformed in some way into a set of features, and the features are used by a decision procedure to assign the event to one category or another ('classification') or to a certain place along an ordination axis ('ordination' in ecology, 'scaling' in psychology). Classification and ordination are clearly processes of information reduction. Machine Learning, as a discipline within AI, is one way of approaching the problem of pattern recognition.

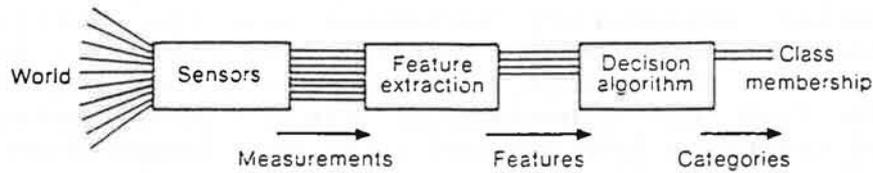


Figure 3 : Pattern recognition (Forsyth & Rada, 1986)

(ii) Machine Learning

Machine Learning is the key to machine intelligence, just as human learning is the key to human intelligence (Forsyth & Rada, 1986). It is the process whereby machines increase their knowledge or improve their skill. When a computer system improves its performance at a given task over time, without re-programming, it can be said that it has learned something. An underlying conviction of many researchers in machine learning is that learning is a prerequisite to any form of intelligence, and therefore it is the core of AI.

All systems designed to modify and improve their performance share certain important common features. Fig. 4. sketches the four major components of a typical learning system, which essentially is a pattern recognizer which learns to associate input descriptions with output categories.

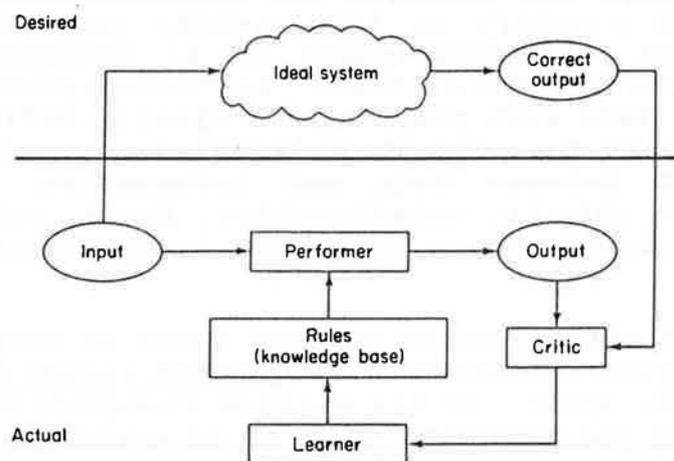


Figure 4 : A framework for learning. (Forsyth, 1989)

The Critic compares the actual with the desired output. In order to do so, there must be an 'ideal system', against which the system's behaviour is measured. In practice this may be a human expert or teacher.

The Learner is responsible for amending the knowledge base to correct erroneous performance. Different learning strategies have been developed.

The Rules are the data structures that encode the system's current level of expertise. They guide the activity of the performance module. The crucial point is whether they can be amended. Real learning systems have in stead of a 'read only' knowledge base rules that constitute a programmable-erasable knowledge base.

Finally, the Performer is the part of the system that carries out the task, using the rules. When the rules are updated, the behaviour of the system changes.

With respect to the possibility of modifying the rules, learning systems can be classified into two main categories : the black box learning systems, also called the geometric or statistical approach, and the structural learning systems (Forsyth & Rada, 1986 ; Jain, 1987).

(a) A "black box" learning system is a system completely specified by its input-output behaviour. Being able to look inside the black box is not one of the design goals, therefore it is not possible changing internal rules. It does not matter whether the behaviour is realized electronically, hydraulically, mechanically, or by brains. Black-box methods share two distinguishing features : (1) a mathematical bias and (2) a 'write-only' description language. The mathematical bias means that they tend to employ well-established procedures from the realms of statistics and control theory. Partly as a consequence of this, the output of the system is opaque. It may calculate a covariance matrix or optimize a set of coefficients, but even a mathematically sophisticated person cannot inspect knowledge in this format and readily determine what the system has learned. This is what is meant by saying that black-box learning systems have a 'write-only' knowledge base. It is computable, but not intelligible. To extract 'knowledge' tables and figures should be constructed and interpreted by experts.

Systems of this group use a description language in which input patterns are presented as feature vectors, always numeric and sometimes binary. Thus an input example is a vector of numbers. If there are p features, then this vector defines a point in p -dimensional space. Various mathematical/geometric terms are used to describe regions in this abstract space. Many of the methods attempt to partition the feature space so that clusters of similar examples are grouped or examples are arranged along a concrete or abstract gradient. These jobs are known as cluster analyses or ordinations.

Learning systems of the black-box type fall into two main groups, on the basis of how they store their knowledge. There are those that adjust the parameters of coefficients of a discriminant function until it is optimal, or at least satisfactory, according to predefined criteria ; and there are those which perform what amounts to an indexing operation : they seek to construct little boxes in feature space such that each box contains only (or mostly) one kind of pattern. Borrowing the terminology of SAMUEL (1967), we say that systems of the second type construct 'signature tables'.

Tree-ring data are often analysed by time-series analysis, by response function analysis or by canonical ordination methods, which could be classified as 'black-box' methods.

The classic book 'Tree Rings and Climate' by Fritts (1976) provided a superb introduction to the science of dendrochronology and an in-depth description of techniques useful for extracting

climatic information from tree-rings. More recently a complete overview of the 'Methods of Dendrochronology and the Applications in the Environmental Sciences' is edited by Cook & Kairiukstis (1990) with excellent contributions of for instance Guiot (1990), Visser & Molenaar (1990), Eckstein (1990),...

Such procedures of the black box type have proved to be useful in a number of pattern-recognition tasks, but the knowledge they acquire tends to be rather opaque. Only the effectiveness of the system matters. Though a component of learning, these methods are not learning in the sense of improving performance at an assigned task.

(b) In contrast, structural learning systems are intended to generate knowledge that is humanly comprehensible as well as it is accessible to machines.

Structural pattern recognition is intuitively appealing. The main advantage of the structural approach over the geometric-black-box approach is that, in addition to classification, it also provides a description of how the given pattern is constructed from the primitives (smallest pieces of information for which further dividing is not relevant). This paradigm has been used in situations where the patterns have a definite structure which can be captured in a set of rules.

(iii) Paradigms for Machine Learning

Four major paradigms (learning strategies) focus the efforts of researchers nowadays (Hellinck & Naydenova, 1990).

(a) Inductive learning. Induction within machine learning mainly involves answering the question 'in what ways can a machine develop general rules from specific examples, and how reliable are those rules in practice?'. The principle is that the expert supplies a set of domain examples of different types of decisions, called a training set, together with the attributes which describe the examples, and values he/she assigns to those attributes. From the training set, a computer program using an inductive algorithm induces a set of rules, which are often constructed in the form of a decision tree.

(b) Analytic learning. Another more recently developed but also very widely used paradigm is based on analytical learning from few examples (sometimes even a single one) within a rich underlying domain theory. The methods that are used in the process are deductive rather than inductive and past problem-solving experience is fully utilized to formulate control rules that enable more efficient application of the available domain knowledge.

(c) Connectionist learning methods. Connectionist learning systems (neural networks) make use of parallel computation in networks of interconnected elements (originally designed as models for the human brain). The main characteristic of the connectionist learning systems is that they discriminate between equivalence classes of patterns from an input domain in a holistic manner.

(d) Genetic or evolutionary algorithms ('classifier systems'). In any evolutionary learning scheme a population of structures (mostly rules) are treated to generate new structures in ways that are explicitly designed to simulate the main

attributes of biological reproduction like cross-over, inversion and mutation. Selection of rules which survive the longest and have greatest likelihood of 'breeding' depends on their performance at the task in hand. This principle is analogous to the survival of the fittest phenomena and natural selection.

(iv) The PC/BEAGLE program

One of the most promising methods currently being investigated in ecology is that of inducing sets of rules from a data set, which consists in our case of as well tree ring variables as environmental variables, by using an evolutionary algorithm : BEAGLE (Biologic Evolutionary Algorithm Generating Logical Expressions). These rules can be very helpful in understanding the major characteristics of systems of trees by classifying objects or by predicting scores on one variable using the other variables in the data set.

The BEAGLE system is a commercially available software package employing evolutionary rule induction. It consists of six main modules which are generally run in sequence (Forsyth, 1989):

SEED Selectively Extracts Example Data
 ROOT Rule-Oriented Optimization Tester
 HERB Heuristic Evolutionary Rule Breeder
 STEM Signature Table Evaluation Module
 LEAF Logical Evaluator And Forecaster
 PLUM Procedural Language Utility Maker

A diagram of how they link together is shown in Fig. 5.

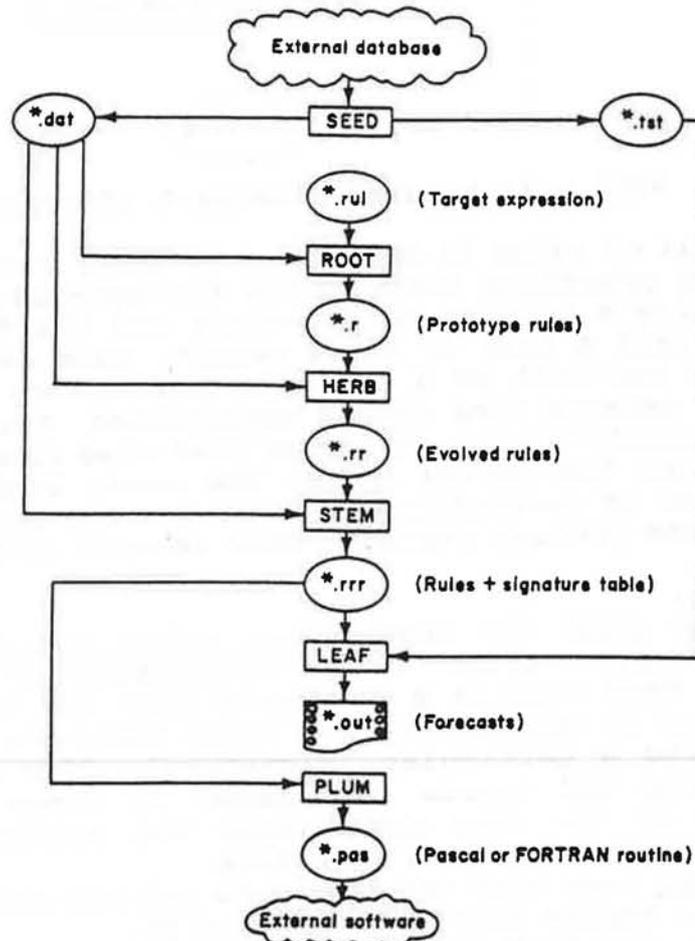


Figure 5 : The linkage of the six main components of BEAGLE (Forsyth, 1989)

(a) SEED is a simple data extraction program. It interfaces BEAGLE to external databases. It can split databases in two (training and test data sets) and it can append leading and/or lagging variables for time-series analyses.

(b) In the ROOT module an initial population of random rules for the breeding process is created. In this module the target expression, which is a logical or numerical expression describing what you want to predict, is formulated.

(c) HERB is the main module: it actually performs the evolutionary process by generating new rules. It takes a datafile from SEED, a tag-file (describing the variables) and an initial rule file (from ROOT) as input. It produces a new rule file as output.

The main program flowchart is shown in fig. 6.

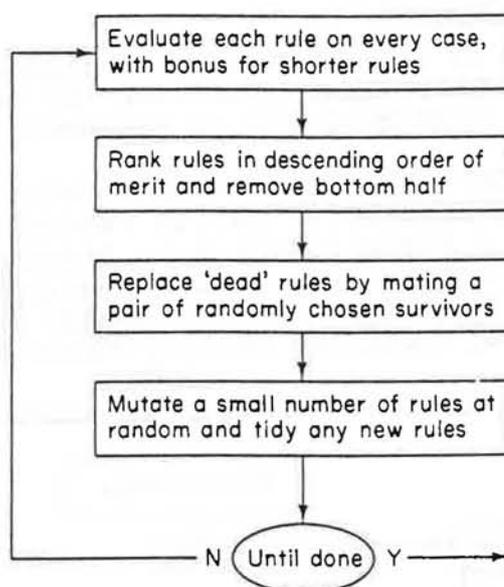


Figure 6 : HERB main program flowchart (Forsyth, 1989)

The survival of rules is determined by their rank order of merit. The scoring procedure rests on the Chi-squared statistics. Each rule can give a true or false result and the target expression can also yield a true or false result. Thus each joint outcome falls into one cell of a fourfold contingency table. The more this table departs from chance expectation, the better i.e. the more effectively the target can be predicted from the rule value. Herb contains two nested loops. The inner loop goes through a given number of generations and retains the best rule found so far. Then the process repeats, thus several rules are generated for output.

(d) In STEM the independent rules are combined into a signature table, which is a forecasting or classification procedure. Each rule is a predicate that can be in one of two states - true or false. With four rules there are 16 combinations each defining a particular 'signature'. STEM re-examines the training data and counts the number of times each signature occurs and at the same accumulates the average value of the target expression for each signature.

(e) LEAF runs over the test data set and estimates what the value of the target expression should be.

(f) Finally PLUM translates a BEAGLE rule file in a Pascal or Fortran subroutine.

On repeated occasions it is showed that BEAGLE is extremely useful for anyone who has to try to predict scores on one variable using the other variables in the data set or to classify objects on the basis of existing data.

As an illustration on these two objectives, forecasting and discrimination, two data sets were selected to be analysed by BEAGLE.

A first data set consists of poplar tree-ring widths and meteorological data. Such a data-set could be analysed by traditional statistical packages which provide multiple regression analysis, canonical ordination or principal component analysis.

The second data set, with yearly diameter data of two poplar clones planted from different planting distances, is to try out the discrimination possibilities of BEAGLE. Traditionally, this problem is approached by variance analysis, cluster analysis or discriminant analysis.

RESULTS

Prediction

With the seed module the space-delimited files are converted in comma-delimited ones. Since there are not enough samples to split up the data set, no test set was created with the Wachtebeke data.

Subsequent BEAGLE-analyses of the Wachtebeke data provide a set of rules, concerning the relations between ring widths of the sample trees and the meteorological variables.

With the logical target-expression (temperature > 0), the trees with the strongest temperature signal could be found.

A high quality rule (score of 87.44) is produced by the LEAF-module: tree5 > tree9. When the summer temperature is higher than the average summer temperature of the 1974-1989 period (target true), there is a good chance that the rule is true : the ring width of tree number 5, growing in a planting pattern of 1107 trees per ha will be broader than the ring width of tree number 9, growing in a planting pattern of 642 trees per ha.

The rule quality is expressed in a sequence of numbers :

87.44 7 0 1 8.

The score 87.44 is obtained from the Phi-coefficient, which is related to the Chi-square statistic and gives an index between -1 and +1, measuring the degree of association between two binary variables. This number is scaled to a value between 0 and 100 and a small penalty which increases with the length of the rule is subtracted. It expresses the departure from pure-chance expectation. Any score over 60 can be considered as excellent, although this depends on the application. The other four numbers are frequencies from a contingency table : 7 cases are true positive, 0 false positive, 1 is false negative and 8 cases are true negatives. In other words : when the summer temperature is higher than the average temperature for the periode 1974-1989 (as is the case for 8 years), the rings of tree number 5 are wider than these of tree number 9, except for one year. When the summer temperature is lower than the average value (also 8 times) the ring width of tree 5 is lower than the radius increment of tree 9.

A high quality rule, revealing a strong precipitation (target : precipitation > 0) signal is :

```
(( tree14 <= tree11 ) & ( tree15 < 0.6114145 ))
86.44 7 1 0 8 .
```

When we don't work with standardized values of tree rings and meteorological data (target : precipitation > 336 mm, being the average summer rainfall), two other rules, combined in a signature table were produced by the STEM-module :

```
precipitation > 336
((tree10 + tree6) < 1,24
85.35 7 1 0 6
(tree1 >= (tree8 - 0.312)
85.35 7 1 0 6
0.5 14
00 0 6 0.0714 0
01 0 0 0.5
10 0 0 0.5
11 7.0 8 0.83 87.5
```

Such a signature table gives four possible combinations (each of the two rules can be true or false). Each of these is a 'fingerprint' or 'signature' that identifies a particular pattern in the data.

STEM gives a line between the rules and the signature table with two numbers : the proportion of cases where rainfall during the growing season is higher than 336 mm and the number of cases in the set. The first column shows the status of the two rules (e.g. 01 meaning that Rule 1 is false and Rule 2 is true). The second item is the number of positive examples (i.e. cases where the target was true). The third item is the count of cases which fell into this rule combination group. The fourth item is the estimate of the target value. The final item is the percentage of cases in this signature group for which the target was true. So the last row of the table states that there were 8 cases where both rules were true, of which 7 or 87.5 % for which summer rainfall was higher than the average. In future cases where both rules are true, the probability that this corresponds with a summer rainfall higher than 336 mm will be estimated as 0.83.

Good rules for insolation (target : insolation > 0) appears to be : ((tree2 >= -0.3563512) <= (tree5 < tree4)), (98.25 10 0 0 6) and ((tree14 >= 0.7885853) <> (tree3 <= -0.3459500)), (98.25 10 0 0 6).

To understand such rules it is important to know that BEAGLE is able to intermix numerical and logical values. If it needs a logical value but is given a numeric one, it evaluates as follows: $x > 0$ corresponds to 'true', $x \leq 0$ corresponds to 'false', 'true' corresponds to 1.0 and 'false' corresponds to 0.0.

With an input of a numeric target expression :

```
(( (tree12 & tree13) & (tree14 & tree15)) - 0.0000)$, it
is possible to find out which meteorological variables or which
combination of meteorological variables are decisive for a
prediction of the radius increment of the poplar trees, growing
in a wide planting pattern. This is expressed in following
signature table :
```

```

! (( amount of sunshine > -0.3307000) | temperature)
$      5      0      1      10
() $ -- from stanmete.dat on 25/01/1991 at 20:30:55
      0.3750      16
00      1.0000      11      0.1146      0.0909
01      5.0000      5      0.8958      1.0000

```

The "!"-mark expresses a logical negation (NOT), the "|" -mark a logical disjunction (OR). From this rule it is possible to conclude that poplar trees in thin stands (planting distance higher than 6 m) show strong signals for lower summer temperatures. The evapotranspiration process might be the growth-limiting factor for these seventeen year old trees. This is an affirmation of an inference based on a redundancy analysis of the same data (Beeckman, 1991).

The meteorological signal of the other trees, growing in denser circumstances, is different, but no high quality rule which makes sense could be found.

With target expressions concerning the maximum growth area it can be affirmed, without plotting graphs or processing growth functions, that poplar trees in narrow planting patterns grow relatively fast the first years after planting, but the growth is slowed down after a couple of years. From that moment relative growth of trees from thinner stands shows a take-off phase. For instance, a target expression (maximum growing area < 10 m²) generates high quality rules (ring width 1978 <= ring width 1976) and (0,55 >= ring width 1981) or a target expression (maximum growing area > 20 m²), gives (ring width 1975 <= ring width 1982).

Discrimination

Based on a t-statistic the hypothesis that there is no general difference between stem diameters of UNAL trees and BEAUPRE trees of the Kaulille experimental field can't be rejected, at alpha = 0.05. Nevertheless, it is possible to discriminate among the two clones with some high quality rules combining the character variables "circle" and "year", as is illustrated by the output of a BEAGLE-session, with the target expression:

```
( UNAL < BEAUPRE).
```

```
$
```

```
Output from HERB :
```

```
(( year <= 1984.0000) <= ( circle > ( year - 1975.0000)))
$      86.50      66      1      2      15.
```

```
Output from STEM :
```

```
(( year <= 1984.0000) <= ( circle > ( year - 1975.0000)))
$      66      1      2      15
() $ -- from diamka.dat on 22/01/1991 at 14:42:01
      0.0000      84
0      2.0000      17      0.1111      11.7647
1      66.0000      67      0.9706      98.5075

```

Simple statistics on data-set -- diamka.dat

Variable		min.	mean	max.	
1	circle	1.0000	6.9524	13.0000	C
2	year	1981.0000	1985.2857	1989.0000	C
3	UNAL	43.2963	169.4576	300.5380	N
4	BEAUPRE	50.1111	179.2828	335.3330	N

Number of samples = 84
Mean target value = 0.8095
Date : 22/01/1991

20 % of the cases are saved in a test data set, to evaluate the rules by the LEAF-module. The LEAF report, sorted on estimated values of the target expression, is like this :

Rank	Actual	Estimate		Circle	year
1	1.00	0.9706	+	8	1982
2	1.00	0.9706	+	10	1982
3	1.00	0.9706	+	11	1982
4	1.00	0.9706	+	10	1984
5	1.00	0.9706	+	2	1985
6	1.00	0.9706	+	9	1986
7	1.00	0.9706	+	6	1987
8	1.00	0.9706	+	9	1987
9	1.00	0.9706	+	10	1987
10	1.00	0.9706	+	12	1987
11	1.00	0.9706	+	2	1988
12	1.00	0.9706	+	8	1988
13	1.00	0.9706	+	11	1988
14	1.00	0.9706	+	1	1989
15	1.00	0.9706	+	13	1989
16	0.00	0.1111	+	2	1981
17	0.00	0.1111	+	2	1982
18	0.00	0.1111	+	4	1982
19	0.00	0.1111	+	6	1982
20	0.00	0.1111	+	8	1984

Crude success rate: 100.0000%
Mean target value = 0.7500

Success rate in non-queried groups = 99.9999% [20 cases].
Av. target value for all-YES group = 1.0000 [15 cases].
Av. target value for All-NO group = 0.0000 [5 cases].

Data-file was : diamka.tst
Rule-file was : diamka.rrr

Rule-set being used (with logical Target Expression):

--> (UNAL < BEAUPRE)

Rule No. 1 :
((year <= 1984.0000) <= (circle > (year - 1975.0000)))
KEY:
=====

?? queries an estimate based on a small signature group, based on less than 15 examples.
+ indicates a "correct" decision :
i.e. estimate >= 0.5 and TRUE target;
or estimate < 0.5 and FALSE target.

HEADINGS:

=====

RANK Ranking sorted by Beagle estimate.
 ACTUAL Actual value of target expression.
 ESTIMATE Beagle's estimate of the target value,
 derived from the signature table.

From the rule it can be concluded that UNAL-trees are smaller than BEAUPRE-trees in following cases : up to and including 1984 (first part of the rule true or = 1), when the number of the circle is bigger than (year - 1975) (second part of the rule true or =1) ; and from 1985 onwards for all the circles.

DISCUSSION

BEAGLE gives evidence to be a very interesting software package with which it is possible to process ecological data. It is especially useful in data sets where the concepts time and complexity are involved, like in dendrochronology. It is laborious and in many cases impossible to handle such datasets with traditional packages, all the more because of a priori assumptions which should be fulfilled, the difficulties with the evaluation of the significance, the lack of detailed descriptions, the difficulties to interpret the output,...

BEAGLE at the other hand is easy to use, does what it claims and is really quit inexpensive (Rowley, 1990). Anyone who suspects that their data may just contain some interesting rules would be well advised to spend an hour or two letting PC/BEAGLE run through the data. Nevertheless, PC/BEAGLE rules should never be blindly accepted, at the very least the rules produced should help in understanding the data. Perhaps the most difficult part is interpreting these rules and attempting to work out whether they make sense and are in the line of the user's expectations. This is not particularly a problem relating to the package. It is simply the difficulty of understanding a mixture of parametric and non-parametric statistics with some Bayesian statistics thrown in too.

A minor disadvantage is that the output files don't include the warnings like 'too few variables to analyse' or 'not enough samples'.

Anyhow it was possible to work out some valuable rules to predict poplar tree-ring widths from meteorological variables or to discriminate among Beaupré and Unal.

ACKNOWLEDGEMENTS

We would like to express our thankfulness towards ir. Norbert Martens, state forester (Dienst Groen, Waters en Bossen) who put forest grounds at the disposal of the university for an experimental poplar plantation and who measured the Kaulille poplars from 1986. Likewise we are grateful towards the still not deceased Laboratories of Plant Ecology and Wood Technology (included the Xylindus team) who sampled the Wachtebeke plot in a hopefully not terminal hand in hand action.

REFERENCES

BARR, A. & FEIGENBAUM, E.A.

BEECKMAN, H. (1989). Mathematische verwerking van fyto-sociologische gegevens. Groene Band 76, 1-36.

BEECKMAN, H. & LEMEURE, R. (1989). Ecology, science of complexity and interactions. Flanders Technology International 1989 Project Sheet Interface RUG 9.

BEECKMAN, H. (1991). Redundancy Analysis of tree-rings and meteorological data in a Nelder design poplar plantation. LUNDQUA Report, Lund, Sweden (in press).

BODEN, M.A. (1977). Artificial Intelligence and Natural Man. The Harvester Press, Hassocks. 537 pp.

BREUKER, J.A. & WIELINGA, B.J. (1983). Analysis techniques for knowledge based systems. Part 2: methods for knowledge acquisition. Report 1.2, Esprit Project 12, University of Amsterdam.

COOK, E. & KAIRIUKSTIS, L. (1990). Methods of dendrochronology. Applications in the Environmental Sciences. L.A. Kluwer Academic Publishers. 394 pp.

ECKSTEIN, D. (1990). Tree-ring/Environment Interactions and their Assessment. In : Methods of dendrochronology. Applications in the Environmental Sciences. Eds. Cook & Kairiukstis. L.A. Kluwer Academic Publishers. 270-276.

FORSYTH, R. (1987). PC/BEAGLE User Guide. Warm Boot Ltd. 55 pp.

FORSYTH, R. (1989). The logic of induction. In : Machine Learning. Principles and Techniques. Ed. : R. Forsyth. Chapman and Hall Computing. p. 1-21.

FORSYTH, R. (1989). The evolution of intelligence. In : Machine Learning. Principles and Techniques. Ed. : R. Forsyth. Chapman and Hall Computing. p.65-82.

FORSYTH, R. & RADA, R. (1986). Machine learning. Applications in expert systems and information retrieval. Ellis Horwood Ltd. Chichester. 277 pp.

FRITTS, H. (1976). Tree Rings and Climate. Academic Press London.

GUIOT, J. (1990). Methods of calibration. In : Methods of dendrochronology. Applications in the Environmental Sciences. Eds. Cook & Kairiukstis. L.A. Kluwer Academic Publishers. 165-177.

HELLINCK, W. & NAYDENOVA, Z. (1990). Automatic knowledge acquisition tools. BIKIT Library Bulletin, Vol. 5, nr. 4.

JAIN, A.K. (1987). Advances in statistical pattern recognition. NATO ASI Series Vol. F30. In : Pattern Recognition, theory and applications. Eds. De Vijver P.A. & Kittler, J. Springer Verlag. Berlin Heidelberg.

LEMEUR, R., Goossens, R. & Schalck, J. (1976). Organization and spatial arrangement of the experimental field at Wachtebeke (Belgium). FKFO Puyenbroeck Populieren Ecosysteem Project, Rapport Nr.2. 38 pp.

NELDER, R. (1962). New kind of systematic designs for spacing experiments. *Biometrics*, 18, 3.

ODUM, E.P. (1971). *Fundamentals of Ecology*. Saunders Philadelphia. 571 pp.

d'OULTREMONT, Comte & STEENACKERS, V. (1972). A propos des nouvelles variétés de peupliers inscrites au contrôle variétal. *Bull. Soc. Roy. For. de Belgique*, 80, 3, pp.110-135.

PRIGOGYNE, I. & STENGERS, I. (1985). *Order out of Chaos*. Fontana Paperbacks, London, 350 pp.

ROWLEY, D.T. (1990). PC/BEAGLE Software Review. *Expert Systems*, Vol.7, No. 1.

SAMUEL, A. (1967). Some studies in machine learning using the game of checkers, Part II. *IBM Journal of Research and Development*, 11.

SCHALCK, J. (1982). Seminar on Biomass Production with fast growing Poplars. Introduction. Laboratories of Plant Ecology and Wood Technology, State University of Ghent.

VISSER, H. & MOLENAAR (1990). Detecting Time-Dependent Climatic Responses in Tree-Rings using the Kalman Filter. *In* : *Methods of dendrochronology. Applications in the Environmental Sciences*. Eds. Cook & Kairiukstis. L.A. Kluwer Academic Publishers. 270-276.

